Einführung in die Stochastik
Übungsblatt 13

TECHNISCHE UNIVERSITÄT DARMSTADT

Fachbereich Mathematik
Dr. Robert Schlicht
Dr. Mehdi Slassi
Dipl. Math. Andreas Fromkorth

Sommersemester 2010
21. Juli 2010

Aufgabe 13.1 — 4 Punkte

Bei der Shell Jugendstudie 2006 wurden 1231 Mädchen befragt. Dabei gaben 55 Prozent der befragten Mädchen an, das Abitur anzustreben. Leiten Sie aus dem zentralen Grenzwertsatz ein approximatives Konfidenzintervall $[u, 1]$ zum Konfidenzniveau 0.95 für den Anteil der Mädchen her, die das Abitur anstreben. (Hinweis: Die Verteilungsfunktion von $N(0, 1)$ nimmt den Wert 0.95 an der Stelle 1.645 an.)

Lösungsvorschlag
Die Anzahl X der Mädchen, die das Abitur anstreben, ist $B_{n,p}$-verteilt ($n = 1231$) unter P_p, $p \in \Theta = (0, 1)$. Wie beschrieben ist also

$$\sqrt{n}\frac{\frac{1}{n}X - p}{\sqrt{\frac{1}{n}X(1-\frac{1}{n}X)}}$$

approximativ $N(0, 1)$-verteilt. Sei $u_\alpha = \Phi^{-1}(1-\alpha)$ das $(1-\alpha)$-Quantil der Standardnormalverteilung. Dann gilt also

$$P_p\left(\sqrt{n}\frac{\frac{1}{n}X - p}{\sqrt{\frac{1}{n}X(1-\frac{1}{n}X)}} \le u_\alpha\right) \approx 1-\alpha.$$

Auflösen der Bedingung in P_p nach p führt zu folgendem approximativen Konfidenzintervall für p:

$$\left[\frac{1}{n}X - u_\alpha\frac{\sqrt{\frac{1}{n}X(1-\frac{1}{n}X)}}{\sqrt{n}},\ 1\right]$$

Wenn nun für $\frac{1}{n}X$ der Wert 0.55 beobachtet worden ist, hat das Konfidenzintervall im Fall $1-\alpha = 0.95$ den Wert

$$\left[0.55 - 1.645\cdot\frac{\sqrt{0.55\cdot 0.45}}{\sqrt{1231}},\ 1\right] = [0.527,\ 1].$$

Aufgabe 13.2 — 4 Punkte

Die Zufallsvariablen $X_1, \dots, X_n$ seien unabhängig und identisch verteilt mit der Verteilungsfunktion

$$F_\theta(x) = \begin{cases} 1-\theta^2/x^2, & \text{falls } x \ge \theta, \\ 0, & \text{sonst.} \end{cases}$$

wobei der Parameter $\theta > 0$ unbekannt ist.
Bestimmen Sie den Maximum-Likelihood-Schätzer T_n für $\tau(\theta) = \theta$.

Lösungsvorschlag
Differenzieren von F_θ in $[\theta, \infty)$ und $(-\infty, \theta)$ liefert eine Dichte

$$f_\theta(x) = \begin{cases} \frac{2\theta^2}{x^3}, & \text{falls } x \ge \theta, \\ 0, & \text{sonst.} \end{cases}$$

Die Likelihood ist gegeben durch

$$L(\theta) = \prod_{i=i}^{n} f_\theta(X_i) = \begin{cases} 2^n \theta^{2n} \prod_{i=1}^{n} X_i^{-3}, & \text{falls } X_i \geq \theta > 0 \text{ für alle } i = 1, \dots, n, \\ 0, & \text{sonst.} \end{cases}$$

Weil θ^{2n} in θ monoton wächst, wird das Maximum angenommen für das größte θ, für das die Bedingung $\theta \leq X_i$ für alle $i = 1, \dots, n$ noch erfüllt ist, d. h. für

$$\theta = \min\{X_1, \dots, X_n\}.$$

Also ist $T_n = \min\{X_1, \dots, X_n\}$ der Maximum-Likelihood-Schätzer für θ.

Aufgabe 13.3 — 4 Punkte

Bei der Größenmessung in einer Gruppe von 9 Personen ergaben sich folgende Körpergrößen [in cm]:

184.2, 182.6, 185.3, 184.5, 186.2, 183.9, 185.0, 187.1, 184.4.

Diese Messwerte werden als Realisationen der Zufallsvariablen $X_1, \dots, X_9$ angenommen, die unabhängig und identisch $N(\mu, \sigma^2)$-verteilt seien.

(a) Geben Sie ein Konfidenzintervall zum Konfidenzniveau 0.99 für den Erwartungswert μ an, falls die Standardabweichung bekannt ist und $\sigma = 2.4$ [cm] beträgt. (Hinweis: Die Verteilungsfunktion von $N(0,1)$ nimmt den Wert 0.995 an der Stelle 2.576 an.)

(b) Welches Konfidenzintervall ergibt sich in (a) für dasselbe Konfidenzniveau, falls die Standardabweichung als unbekannt angenommen wird? (Hinweis: Die Verteilungsfunktion der t-Verteilung mit 8 Freiheitsgraden nimmt den Wert 0.995 an der Stelle 3.355 an.)

Lösungsvorschlag

(a) $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma}$ ist unter jedem Parameter $N(0,1)$-verteilt (d. h. ein Pivot für μ), liegt also mit Wahrscheinlichkeit 0.99 im Intervall $[-u_{0.995}, u_{0.995}]$, wobei $u_{0.995} = \Phi^{-1}(0.995) = 2.576$ das 0.995-Quantil von $N(0,1)$ ist. Auflösen der Bedingung nach μ führt zu dem Konfidenzintervall

$$\left[\frac{1}{n} \sum_{i=1}^{n} X_i - u_{0.995} \frac{\sigma}{\sqrt{n}}, \quad \frac{1}{n} \sum_{i=1}^{n} X_i + u_{0.995} \frac{\sigma}{\sqrt{n}} \right]$$

für μ. Für die beobachteten Daten und $n = 9$ ergibt sich für $\frac{1}{n} \sum_{i=1}^{n} X_i$ der Wert 184.8 und für das Intervall der Wert

$$[182.739, 186.861].$$

(b) Nach Vorlesung hat $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mu}{\sqrt{\hat{v}}}$ mit $\hat{v} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \frac{1}{n} \sum_{j=1}^{n} X_j)^2$ (Stichprobenvarianz) unter allen Parametern (μ, σ^2) die t-Verteilung mit 8 Freiheitsgraden. Mit dem 0.995-Quantil $t_{8;0.995} = 3.355$ dieser Verteilung erhält man wie in (a) das Konfidenzintervall

$$\left[\frac{1}{n} \sum_{i=1}^{n} X_i - t_{8;0.995} \frac{\sqrt{\hat{v}}}{\sqrt{n}}, \quad \frac{1}{n} \sum_{i=1}^{n} X_i + t_{8;0.995} \frac{\sqrt{\hat{v}}}{\sqrt{n}} \right].$$

Für die gegebenen Daten nimmt $\frac{1}{n} \sum_{i=1}^{n} X_i$ den Wert 184.8 an, $\hat{v}$ den Wert 1.725 und das Konfidenzintervall den Wert

$$[183.331, 186.269].$$

Aufgabe 13.4 — 4 Punkte

Ein Autohersteller behauptet, dass der Benzinverbrauch für einen neuentwickelten Typ im Mittel 6 Liter pro 100km beträgt. Dabei kann er davon ausgehen, dass der Verbrauch normalverteilt ist mit der Standardabweichung $\sigma = 0.3$.

(a) Eine Verbraucherzentrale vermutet, dass der Hersteller einen zu niedrigen Mittelwert μ angegeben hat. Sie überprüft deshalb 20 Autos des neuen Typs auf ihren Verbrauch und berechnet aus diesen Werten das arithmetische Mittel $\bar{x} = 6.1$. Kann man hiermit die Behauptung des Herstellers widerlegen? Führen Sie einen Test zum Niveau $\alpha = 0.05$ durch.

(b) Eine Autozeitschrift bekommt von 152 Käufern des neuen Typs Beschwerden über den zu hohen Verbrauch zugesandt. Sie errechnet aus diesen Werten das arithmetische Mittel $\bar{x} = 7.3$. Kann man hiermit die Behauptung des Herstellers widerlegen?

Lösungsvorschlag

Wir nehmen an $X_1, \ldots, X_n$ unabhängig $N(\mu, \sigma^2)$-verteilt unter P_μ, $\mu \in \Theta = \mathbb{R}$. Wir führen einen Gauß-Test zum Niveau $\alpha = 0.05$ durch. Es soll die Nullhypothese $H_0 : \mu \le 6$ gegen die Alternative $H_1 : \mu > 6$ getestet werden, d. h. wir betrachten

$$\Theta_0 = \{\mu \in \Theta : \mu \le 6\} \quad \text{und} \quad \Theta_1 = \{\mu \in \Theta : \mu > 6\}.$$

Gemäß Vorlesung ist der Test gegeben durch

$$\varphi = 1_{\{T(X_1,\ldots,X_n) > \Phi^{-1}(1-\alpha)\}}$$

mit der Teststatistik

$$T(X_1, \ldots, X_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - 6}{\sigma} = \frac{\sqrt{n}}{\sigma}(\bar{x} - 6).$$

(a) Für den gegebenen Wert von $\bar{x}$ nimmt $T(X_1, \ldots, X_n)$ den Wert $\frac{\sqrt{20}}{0.3}(6.1 - 6) = 1.49$ an. Wegen $\Phi^{-1}(1-\alpha) = 1.645 \ge 1.49$ kann die Nullhypothese also nicht abgelehnt werden. Damit ist die Angabe des Herstellers bzgl. des gegebenen Niveaus $\alpha = 0.05$ nicht wiederlegt.

(b) Für den gegebenen Wert von $\bar{x}$ nimmt $T(X_1, \ldots, X_n)$ hier den Wert $\frac{\sqrt{152}}{0.3}(7.3 - 6) = 53.42 > 1.65$ an. Die Nullhypothese kann also abgelehnt werden. Man könnte das Testergebnis trotzdem anzweifeln wegen einem sampling bias bei der Fahrzeugauswahl. Es haben sich eventuell nur die Leute gemeldet, die einen erhöhten Verbrauchswert festgestellt haben. Dies kann aus Übertreibung, Defekten des Fahrzeuges, Fahrweise des Fahrzeughalters und ähnlichem resultieren.